JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 63, 177–198 (1978)

Differential Gradient Methods

C. A. BOTSARIS[1]

National Research Institute for Mathematical Sciences of the Council for Scientific and Industrial Research, P.O. Box 395, Pretoria 0001, South Africa

Submitted by E. Stanley Lee

A class of recently developed differential descent methods for function minimization is presented and discussed, and a number of algorithms are derived which minimize a quadratic function in a finite number of steps and rapidly minimize general functions. The main characteristics of our algorithms are that a more general curvilinear search path is used instead of a ray and that the eigensystem of the Hessian matrix is associated with the function minimization problem. The curvilinear search paths are obtained by solving certain initial-value systems of differential equations, which also suggest the development of modifications of known numerical integration techniques for use in function minimization. Results obtained on testing the algorithms on a number of test functions are also given and possible areas for future research indicated.

1. INTRODUCTION

The problem with which we are concerned may be stated as:

$$\text{minimize } f(x),\ f: R^n \to R^1,\ f \in C^2. \tag{1.1}$$

As may be proved, x^* is a local minimizer of $f(x)$ if

$$g(x^*) = 0, \qquad H(x^*) > 0 \tag{1.2}$$

where $g(x)$ is the gradient vector, $g^T(x) = [\ldots, \partial f(x)/\partial x_i, \ldots]$, and $H(x)$ is the Hessian matrix, $H(x) = [\partial^2 f(x)/\partial x_i\, \partial x_j]$, of $f(x)$.

All existing methods for minimization generate a sequence $\{x^k\}$ of approximate minimizers via the iteration formula (see [1])

$$x^{k+1} = x^k + t^k d^k \tag{1.3}$$

where d^k is a properly defined descent direction and t^k the step size so chosen that

$$f(x^{k+1}) - f(x^k) < 0. \tag{1.4}$$

[1] Present address: Department of Mathematics, School of Mathematics and Physics, Aristotle's University of Thessaloniki, Thessaloniki, Greece.

0022-247X/78/0631-0177$02.00/0

A very broad, and perhaps the most important class of methods for unconstrained minimization is that consisting of the co-called gradient methods which set

$$d^k = -Q^{-1}(x^k)\, g(x^k) \tag{1.5}$$

where $Q(x)$ is a positive definite symmetric matrix. As may be proved, [2], equation (1.5) defines the steepest descent direction for $f(x)$ at $x = x^k$, under the general elliptic norm $\| d \| = +(d^T Q d)^{\frac{1}{2}}$.

Substituting the identity matrix, I for $Q(x)$ in (1.5) and using (1.3), we obtain the iteration formula

$$x^{k+1} = x^k - t^k g(x^k) \tag{1.6}$$

which may be recognized as the oldest known minimization method, namely that of steepest descent. Newton's method is obtained by letting $Q(x) = H(x)$, in which case (1.3) becomes

$$x^{k+1} = x^k - t^k H^{-1}(x^k)\, g(x^k) \tag{1.7}$$

whereas variable metric methods, [3], [4], set

$$x^{k+1} = x^k - t^k \eta^{-1}(x^k)\, g(x^k) \tag{1.8}$$

where $\eta^{-1}(x^k)$ is an approximation to $H^{-1}(x^k)$ and is updated according to

$$\eta^{-1}(x^{k+1}) = \eta^{-1}(x^k) + \frac{\Delta x_k (\Delta x_k)^T}{(\Delta x_k)^T \Delta g_k} - \frac{\eta^{-1}(x^k)\, \Delta g_k (\Delta g_k)^T\, \eta^{-1}(x^k)}{(\Delta g_k)^T\, \eta^{-1}(x^k)\, \Delta g_k} \tag{1.9}$$

(Davidon-Fletcher-Powell, [5]) or

$$\eta^{-1}(x^{k+1}) = \left(I - \frac{\Delta x_k (\Delta g_k)^T}{(\Delta x_k)^T \Delta g_k}\right) \eta^{-1}(x_k) \left(I - \frac{\Delta g_k (\Delta x_k)^T}{(\Delta x_k)^T \Delta g_k}\right) + \frac{\Delta x_k (\Delta x_k)^T}{(\Delta x_k)^T \Delta g_k} \tag{1.10}$$

(Fletcher, [6]) where $\Delta x_k = x^{k+1} - x^k$, $\Delta g_k = g(x^{k+1}) - g(x^k)$, $\eta^{-1}(x^0) = I$.

It has, however, been suggested by various authors in the past [7][8], that the rays along which successive iterates are obtained when a function is minimized by means of an iterative method, should be replaced by more general curvilinear paths. Our algorithms are of this type in that they replace the iteration formula $x^{k+1} = x^k + t^k d^k$ by

$$x^{k+1} = x^k + p(x^k, t^k) \tag{1.11}$$

where $p(x^k, t)$ is not a ray, but a curve in x-space. These curvilinear search paths are obtained by solving certain systems of differential equations, and a unified way of obtaining them is presented in the following section.

2. Derivation of the Basic Systems of Differential Equations

It is known that for any function $f: R^n \to R^1$ and any given point $x^k \in R^n$ the steepest descent direction d^k is defined as the direction of maximal local decrease of $f(x)$. Moreover, if $f(x)$ is differentiable at x^k, then this direction is that for which $g^T(x^k)d/\| d \|$ takes on its minimum value as a function of $d \neq 0$. Since $g^T(x^k)d$ is a continuous function of d and the set $\{d, \| d \| = 1\}$ is compact, it follows that such a direction always exists, although it is not unique, as it depends on the particular norm $\| \cdot \|$. As we mentioned in the previous section, under the general elliptic norm $\| d \| = +(d^TQd)^{\frac{1}{2}}$ the steepest-descent direction d^k at $x = x^k$ is given by

$$d^k = -Q^{-1}(x^k)\, g(x^k). \tag{2.1}$$

We now generalize the concept of a steepest-descent direction and define a steepest-descent curve at $x^k \in R^n$ as a curve of maximal local decrease of $f(x)$. Based on the concept of a steepest-descent curve, our first-order systems of differential equations are now derived.

2.1. Lemma. *Let $f: R^n \to R^1$ be a continuously differentiable function and let $X = \{x \mid x: R^1 \to R^n, x(0) = x^k\}$ be the set of all differentiable space curves passing through a given point $x^k \in R^n$. Then a steepest-descent curve at x^k obeys the following initial-value system of differential equations*:

$$\begin{aligned} \dot{x}(t) &= -Q^{-1}(x)\, g(x) \\ x(0) &= x^k,\ \dot{x}(0) \neq 0 \end{aligned} \tag{2.2}$$

under the general elliptic norm $\| d \| = +(d^TQ(x)d)^{\frac{1}{2}}$, $Q(x) > 0$.

Proof. It is known from calculus that the change of $f(x)$ at $x = x^k$ along a space curve $x(t)$, $x(0) = x^k$, is given by $df/ds \mid_{s=0}$, where s is a real parameter defined as the distance moved along $x(t)$ and given by

$$s(t) = \int_0^t \| \dot{x}(t) \| \, dt. \tag{2.3}$$

Hence the maximal local decrease of $f(x)$ at x^k results when $df/ds \mid_{s=0}$ assumes its minimum value as a function of s. At this point we recall that

$$\frac{df(x)}{ds} = \frac{df(x)}{dt} \frac{1}{(ds(t)/dt)} = \frac{df(x)}{dt} \frac{1}{\| \dot{x}(t) \|} \tag{2.4}$$

and

$$\frac{df(x)}{dt} = g^T(x)\, \dot{x}(t). \tag{2.5}$$

Therefore

$$\left.\frac{df(x)}{ds}\right|_{s=0} = \left.\frac{g^T(x)\,\dot{x}(t)}{\|\dot{x}(t)\|}\right|_{t=0} = \frac{g^T(x^k)\,\dot{x}(0)}{\|\dot{x}(0)\|}. \tag{2.6}$$

We now have, using the Cauchy-Schwartz inequality,

$$\begin{aligned}
[g^T(x^k)\,\dot{x}(0)]^2 &= [g^T(x^k)\,Q^{-1}(x^k)\,Q^{\frac{1}{2}}(x^k)\,Q^{\frac{1}{2}}(x^k)\,\dot{x}(0)]^2 \\
&\leqslant [g^T(x^k)\,Q^{-1}(x^k)\,Q^{\frac{1}{2}}(x^k)][Q^{\frac{1}{2}}(x^k)\,Q^{-1}(x^k)\,g(x^k)] \\
&\quad \times [\dot{x}^T(0)\,Q^{\frac{1}{2}}(x^k)][Q^{\frac{1}{2}}(x^k)\,\dot{x}(0)] \\
&= \|Q^{-1}(x^k)\,g(x^k)\|^2\,\|\dot{x}(0)\|^2.
\end{aligned} \tag{2.7}$$

Clearly the minimum value of $df(x)/dx\,|_{s=0}$ as a function of $\dot{x}(0) \neq 0$ results when the equality sign holds in (2.7), i.e. when

$$\dot{x}(0) = -Q^{-1}(x^k)\,g(x^k). \tag{2.8}$$

Hence $x(t)$ obeys the following system of differential equations:

$$\begin{aligned}
\dot{x}(t) &= -Q^{-1}(x)\,g(x) \\
x(0) &= x^k,\ \dot{x}(0) \neq 0
\end{aligned} \tag{2.9}$$

and the lemma is proved. ▌

It is now clear that different choices of the matrix $Q(x)$ will lead to different steepest-descent curves. Thus we obtain

$$Q(x) = I \Rightarrow \dot{x}(t) = -g(x) \tag{2.10}$$

$$Q(x) = H(x) \Rightarrow \dot{x}(t) = -H^{-1}(x)\,g(x) \quad x(0) = x^k. \tag{2.11}$$

$$Q(x) = H^{-1}(x) \Rightarrow \dot{x}(t) = -H(x)\,g(x) \tag{2.12}$$

The importance of the above systems of differential equations lies in the fact that if the actual solution curves to them were available, the minimum would be reached in one step for any function. The reason is that the solution curves to the above systems tend asymptotically to a local minimizer, x^*, of $f(x)$. Indeed, referring to system (2.10), the solution curve to it is normal to the countours of $f(x)$, having therefore to tend to x^*. The same argument holds for system (2.12) if instead of $f(x)$ we consider the real functional $\phi(x) = \frac{1}{2}g^T(x)\,g(x)$. Obviously $\phi(x) \geqslant 0$ has a minimum value of zero at x^*, as at this point $g(x^*) = 0$. Note that $\partial\phi(x)/\partial x = H(x)\,g(x)$, and therefore the solution curve to system (2.12), being normal to the countours of $\phi(x)$, tends asymptotically to x^*. For system (2.11) this is not immediately obvious, however, and a different argument is now applied to it.

As we mentioned in the previous section, if x^* is a local minimizer of $f(x)$, then $g(x^*) = 0$. Therefore minimizing $f(x)$ over R^n is equivalent to solving a certain system of non-linear equations:

$$g(x) = 0. \tag{2.13}$$

Such a system may be solved by the method of differentiation with respect to a parameter which imbeds the real parameter t into the original equation. This is done by considering instead of the single mapping $g\colon R^n \to R^n$ an entire family $F\colon R^n \times [0, \infty) \to R^n$ defined by

$$F(x, t) = g(x) - e^{-t}g(x^k). \tag{2.14}$$

As may proved [2], if $H^{-1}(x)$ is bounded, i.e. $\| H^{-1}(x)\| \leqslant b, 0 \leqslant b < +\infty$, then for any fixed $x^k \in R^n$ there exists a unique mapping $x\colon [0, +\infty) \to R^n$, $x(0) = x^k$ such that

$$F(x(t), t) \equiv 0, \qquad \forall t \in [0, +\infty). \tag{2.15}$$

Moreover

$$\dot{x}(t) = -H^{-1}(x)\, g(x). \tag{2.16}$$

Note that because of (2.15)

$$g(x(t)) = e^{-t}g(x^k) \tag{2.17}$$

along $x(t)$. Now let $x^* = \lim_{t\to+\infty} x(t)$. Then from (2.14) we obtain, because of (2.15) and the continuity of $g(x)$,

$$\lim_{t\to+\infty} F(x, t) = \lim_{t\to+\infty} g(x(t)) = g(x^*) = 0.$$

Hence $x(t)$ tends asymptotically to a root of $g(x) = 0$, i.e. to a local minimizer x^* of $f(x)$, provided that $H(x^*)$ is positive semi-definite.

From the previous analysis it follows that if we start at x^k and move, in the direction of decreasing values of $f(x)$, along the solution curve to any of our systems of differential equations, this will bring us directly to the minimum, i.e. the minimum is reached in one step for any function.

Higher-order systems of differential equations may also be derived based on the following lemma.

2.2 LEMMA. *Let $f\colon R^n \to R^1$ be a continuously differentiable function and $x\colon R^1 \to R^n$, $x(0) = x^k$, a differentiable space curve. Let also the first nonzero* *derivative of $f(x)$ along $x(t)$ at $t = 0$ be negative. Then there exists a $\delta > 0$ and sufficiently small such that*

$$f(x) - f(x^k) < 0, \qquad \forall t \in (0, \delta)$$

and consequently $x(t)$ is a descent curve.

Proof. Let

$$\frac{d^q(f(x(0))}{dt^q} < 0,\ 0 < q < +\infty,\ \frac{d^r f(x(0))}{dt^r} = 0,\ r = 1, 2, \ldots, q-1 \qquad (2.18)$$

and let us expand $f(x(t))$: $R^1 \to R^1$ in a Taylor series about $t = 0$. We have

$$\begin{aligned} f(x(t)) &= f(x(0)) + \sum_{j=1}^{r} \frac{1}{j!} \frac{d^j f(x(0))}{dt^j} t^j + \frac{1}{q!} \frac{d^q f(x(0))}{dt^q} t^q \\ &+ \frac{1}{(q+1)!} \frac{d^{(q+1)} f(x(\xi))}{dt^{(q+1)}} t^{(q+1)} \end{aligned} \qquad (2.19)$$

where $\xi \in (0, t)$. Using (2.18) we obtain from (2.19)

$$f(x) - f(x^k) = \frac{1}{q!} \frac{d^q f(x(0))}{dt^q} t^q + \frac{1}{(q+1)!} \frac{d^{(q+1)} f(x(\xi))}{dt^{(q+1)}} t^{(q+1)}. \qquad (2.20)$$

It is now possible to find a $\delta > 0$ such that any $t \in (0, \delta)$ is sufficiently small to make the first term in the expansion (2.20) dominate. Hence since $d^q f(x(0))/dt^q < 0$, we obtain

$$f(x) - f(x^k) < 0, \qquad \forall t \in (0, \delta) \qquad (2.21)$$

thus proving the lemma. ∎

Let us now consider the initial-value system of differential equations

$$\begin{aligned} &x^{[q]}(t) = -Q^{-1}(x)\, g(x) \\ &x^{[r]}(0) = 0,\ r = 1, 2, \ldots, q-1,\ x(0) = x^k. \end{aligned} \qquad (2.22)$$

As may be proved, the solution curve $x(t)$ to the above system satisfies the conditions of the previous lemma, and therefore $x(t)$ is a descent curve in x-space. Note that, in general, $x(t)$ is not a steepest-descent curve as this depends on the higher, up to order q, partial derivatives of $f(x)$.

3. Solution of the Differential Equations

In general it is not easy to solve our systems of differential equations, owing to the non-linearity of $g(x)$. An approximate solution curve may, however, be obtained by expanding $Q^{-1}(x)\,g(x)$ in a Taylor series about x^k. We have:

(i) zero approximation:

$$Q^{-1}(x)\,g(x) \simeq Q^{-1}(x^k)\,g(x^k) \tag{3.1}$$

(ii) linear approximation:

$$Q^{-1}(x)\,g(x) \simeq Q^{-1}(x^k)\,g(x^k) + J(x^k)\,(x - x^k) \tag{3.2}$$

where $J(x)$ is the Jacobian of $Q^{-1}(x)\,g(x)$.

At this point we recall that $Q(x)$ is either the identity matrix I, or $H(x)$ or $H^{-1}(x)$, and that

$$\begin{aligned} J(x) &= \frac{\partial}{\partial x}\,[Q^{-1}(x)\,g(x)] = \frac{\partial}{\partial x}\,[Q^{-1}(x)]\,g(x) + Q^{-1}(x)\,\frac{\partial g(x)}{\partial x} \\ &= \frac{\partial}{\partial x}\,[Q^{-1}(x)]\,g(x) + Q^{-1}(x)\,H(x). \end{aligned} \tag{3.3}$$

Let now $y_i^T(x)$, $i = 1, 2,\ldots, n$, $y_i\colon R^n \to R^n$, be the rows of $Q(x)$. Then $\partial Q^{-1}(x)/\partial x$ is the n-tuple

$$\frac{\partial}{\partial x}\,[Q^{-1}(x)] = \left[\ldots, \frac{\partial y_i(x)}{\partial x}, \ldots\right]. \tag{3.4}$$

Hence

$$J(x) = \left[\ldots, \frac{\partial y_i(x)}{\partial x}, \ldots\right] g(x) + Q^{-1}(x)\,H(x). \tag{3.5}$$

We have three possibilities, as follows.

(a) $Q(x) = I$ *System* $x^{[q]}(t) = -g(x)$, $x^{[r]}(0) = 0$,

$$r = 1,\ldots, q - 1, \qquad x(0) = x^k. \tag{3.6}$$

Then $y_i^T(x) = [0,\ldots, 1,\ldots, 0]$. Therefore $\partial y_i(x)/\partial x = 0$ and (3.5) reduces to

$$J(x) = H(x). \tag{3.7}$$

(b) $Q(x) = H(x)$ *System* $x^{[q]}(t) = -H^{-1}(x)\,g(x)$, $x^{[r]}(0) = 0$,

$$r = 1, 2,\ldots, q - 1, \qquad x(0) = x^k. \tag{3.8}$$

We have

$$J(x) = I + \sum_{i=1}^{n} \left[\frac{\partial f(x)}{\partial x_i}\right] \left[\frac{\partial y_i(x)}{\partial x}\right] \tag{3.9}$$

Note that in general $J(x)$ is not symmetric, and that for a quadratic function

$$J(x) = I, \tag{3.10}$$

as in this case $H(x)$ and hence $H^{-1}(x)$ are constant matrices, independent of x.

(c) $Q(x) = H^{-1}(x)$ *System* $x^{[q]}(t) = -H(x)\,g(x)$, $x^{[r]}(0) = 0$

$$r = 1, 2, \ldots, q - 1, \qquad x(0) = x^k. \tag{3.11}$$

In this case $Q^{-1}(x) = H(x)$ and therefore $y_i(x) = \partial/\partial x[\partial f(x)/\partial x_i]$. Hence

$$J(x) = [H(x)]^2 + \sum_{i=1}^{n} \left[\frac{\partial f(x)}{\partial x_i}\right] \frac{\partial^2}{\partial x^2} \left[\frac{\partial f(x)}{\partial x_i}\right]. \tag{3.12}$$

Note that $J(x)$ is always symmetric, as the Hessian matrix of the functional $\phi(x) = \frac{1}{2}g^T(x)\,g(x)$, and that for a quadratic function

$$J(x) = [H(x)]^2. \tag{3.13}$$

Under the zero approximation (i) we now obtain for systems (3.6), (3.8) and (3.11):

$$x(t) = x^k - \frac{t^q}{q!}\,g(x^k) \tag{3.14}$$

$$x(t) = x^k - \frac{t^q}{q!}\,H^{-1}(x^k)\,g(x^k) \tag{3.15}$$

and

$$x(t) = x^k - \frac{t^q}{q!}\,H(x^k)\,g(x^k) \tag{3.16}$$

respectively. We can easily recognize (3.14) and (3.15) as the steepest-descent method and Newton's method respectivelt, with the change of variable

$$\tau = \frac{t^q}{q!}\,. \tag{3.17}$$

Let us now consider our general system

$$\begin{aligned} &x^{[q]}(t) = -Q^{-1}(x)\,g(x) \\ &x^{[r]}(0) = 0,\ r = 1, 2, \ldots, q - 1,\ x(0) = x^k \end{aligned} \tag{3.18}$$

under the linear approximation (ii). As may be proved

$$x(t) = x^k + [\exp[-t[J(x^k)]^{1/q}] - I]\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k), \qquad q = \text{odd} \quad (3.19)$$

and

$$x(t) = x^k + [\cos[-t[J(x^k)]^{1/q}] - I]\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k), \qquad q = \text{even} \quad (3.20)$$

where $\exp(A)$ and $\cos(A)$ are the matrix exponential and matrix cosine functions of A respectively.

In the following section the properties of the derived solution curves are established and discussed.

4. Properties of the Curvilinear Search Paths

Let

$$P(x^k, t) = \begin{cases} [\exp[-t[J(x^k)]^{1/q}] - I]\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k), \; q = \text{odd} & (4.1) \\ [\cos[-t[J(x^k)]^{1/q}] - I]\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k), \; q = \text{even}. & (4.2) \end{cases}$$

We first note that

$$P(x^k, t) \to \begin{cases} (e^{-t} - 1)\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k) & (4.3) \\ (\text{cost} - 1)\, J^{-1}(x^k)\, Q^{-1}(x^k)\, g(x^k) & (4.4) \end{cases} \quad \text{as } q \to +\infty.$$

Note that in the case of system (3.6) or when the objective function is quadratic, we obtain

$$x(\tau) = x^k - \tau H^{-1}(x^k)\, g(x^k)\; \tau = \begin{cases} 1 - e^{-t} \\ 1 - \text{cost} \end{cases} \quad (4.5)$$

which is Newton's method with a change of variable.

In the sollowing, the curvilinear search paths associated with our first-order systems ($q = 1$) are discussed in more detail.

System

$$\dot{x}(t) = -g(x), \qquad x(0) = x^k \quad (4.6)$$

Solution

$$x(t) = x^k + [\exp(-tH(x^k)) - I]\, H^{-1}(x^k)\, g(x^k). \quad (4.7)$$

This curvilinear search path has been extensively examined in [9] from where we recall the following.

(i) Expanding the matrix exponential function into the eigenspaces of the Hessian matrix we obtain

$$x(t) = x^k + \left[\sum_{i=1}^{n} \frac{\exp(-t\lambda_i(x^k)) - 1}{\lambda_i(x^k)} u_i(x^k)\, u_i^T(x^k)\right] g(x^k) \tag{4.8}$$

where $\lambda_i(x^k)$ are the eigenvalues and $u_i(x^k)$ the associated normalized eigenvectors of $H(x^k)$, $i = 1, 2,..., n$, and as may be seen no matrix inversion is required.

(ii) Equation (4.8) defines always a steepest-descent curve in x-space for $t \geqslant 0$ irrespective of whether or not $H(x^k)$ is positive definite as

$$\frac{\exp(-t\lambda_i(x^k)) - 1}{\lambda_i(x^k)} < 0, \qquad \text{for} \quad \lambda_i(x^k) \neq 0 \tag{4.9}$$

and

$$\frac{\exp(-t\lambda_i(x^k)) - 1}{\lambda_i(x^k)} \to -t \qquad \text{as} \quad \lambda_i(x^k) \to 0. \tag{4.10}$$

Note that when $H(x^k)$ is singular ($\lambda_i(x^k) = 0$ for one or more i, $1 \leqslant i \leqslant n$), Newton's method breaks down, as in this case $H^{-1}(x^k)$ does not exist. Also, if $H(x^k)$ is not positive definite, then Newton's method or variable metric methods may fail, as $H^{-1}(x^k)\, g(x^k)$ or $\eta^{-1}(x^k)\, g(x^k)$ may not be a descent direction.

(iii) Depending on the step size t, (4.7) behaves as the method of steepest descent or Newton's method. Indeed, for small t we have $\exp(-tH(x^k)) \simeq I - tH(x^k)$. Hence $x(t) \simeq x^k - tg(x^k)$, which is the steepest-descent method. Now, for $t \to +\infty$ and provided that $H(x^k)$ is positive definite, we have $\exp(-tH(x^k)) \to 0$ and therefore $x(t) \to x^k - H^{-1}(x^k)\, g(x^k)$, which is the point obtained when Newton's method is used with a step size of unity.

At this point we recall that it has been suggested that a combination of Newton's method and that of steepest-descent should exhibit a performance superior to that of either method alone. Indeed, at a considerable distance from the minimum the steepest-descent method may be superior to that of Newton which has the distinct advantage of quadratic convergence only in the vicinity of the minimum, where the objective function can be approximated reasonably well by a quadratic function. However, in our case such a combination is not necesary, since (4.7) exhibits the same behavior, depending on the step size, as either one of the other methods, and will adapt automatically to suit.

(iv) The space curve defined by equation (4.7) is the actual solution curve to system (4.6) for a quadratic objective function, and therefore for such a function the minimum is reached in one step.

System

$$\dot{x}(t) = -H^{-1}(x)\, g(x), \qquad x(0) = x^k \tag{4.11}$$

Solution

$$x(t) = x^k + [\exp(-tJ(x^k)) - I]\, J^{-1}(x^k)\, H^{-1}(x^k)\, g(x^k). \tag{4.12}$$

We recall that $J(x)$ is the Jacobian of $H^{-1}(x)\, g(x)$ given by

$$J(x) = I + \sum_{i=1}^{n} \left[\frac{\partial f(x)}{\partial x_i} \frac{\partial y_i(x)}{\partial x} \right] \tag{4.13}$$

where $y_i^T(x)$, $i = 1, 2, \ldots, n$, are the rows of $H^{-1}(x)$; and that for a quadratic function

$$J(x) = I. \tag{4.14}$$

Following the same argument as before, we can expand the matrix exponential function into the eigenspaces of $J(x^k)$. Since $J(x)$ is not, in general, symmetric we obtain

$$x(t) = x^k + V(x^k)[\exp(-tM(x^k)) - I]\, M^{-1}(x^k)\, V^{-1}(x^k)\, H^{-1}(x^k)\, g(x^k) \tag{4.15}$$

where $V(x)$ is the modal matrix of $J(x)$, and $M(x)$ its Jordan canonical form. Note that, for small t, $\exp(-tJ(x^k)) \simeq I - tJ(x^k)$ and therefore (4.12) reduces to

$$x(t) \simeq x^k - tH^{-1}(x^k)\, g(x^k) \tag{4.16}$$

which is Newton's method again. Now for a quadratic function we have, because of (4.14),

$$x(t) = x^k + (e^{-t} - 1)\, H^{-1}(x^k)\, g(x^k) \tag{4.17}$$

and letting $t \to +\infty$ we obtain

$$x(t) \to x^k - H^{-1}(x^k)\, g(x^k) = x^*, \tag{4.18}$$

i.e. the minimum is reached on one step. Note also that $df(x(0))/dt = -g^T(x^k)\, H^{-1}(x^k)\, g(x^k)$ and hence (4.12) is a descent curve, according to lemma 2.2, if $H(x^k)$ is positive definite.

System

$$\dot{x}(t) = -H(x)\, g(x), \qquad x(0) = x^k \tag{4.19}$$

Solution

$$x(t) = x^k + [\exp(-tJ(x^k)) - I]\, J^{-1}(x^k)\, H(x^k)\, g(x^k) \tag{4.20}$$

At this point we recall that $x(t)$, as defined by (4.19), is a steepest-descent curve for the functional

$$\phi(x) = \tfrac{1}{2} g^T(x)\, g(x) = \tfrac{1}{2} \| g(x) \|^2 \tag{4.21}$$

under the usual Euclidean norm, and that $J(x)$ is the Hessian matrix of $\phi(x)$ given by

$$J(x) = [H(x)]^2 + \sum_{i=1}^{n} \left[\frac{\partial f(x)}{\partial x_i}\right] \frac{\partial^2}{\partial x^2} \left[\frac{\partial f(x)}{\partial x_i}\right]. \tag{4.22}$$

Note that for $x(t)$ to be a descent curve we must have $g^T(x^k)\, H(x^k)\, g(x^k) > 0$. Again using the expansion of the matrix exponential function into the eigenspaces of $J(x)$, and taking into account that $J(x)$ is symmetric, we obtain

$$x(t) = x^k + V(x^k)[\exp(-tM(x^k)) - I]\, M^{-1}(x^k)\, V^T(x^k)\, H(x^k)\, g(x^k). \tag{4.23}$$

Now for a quadratic function, or near the minimum of a general function, we have

$$J(x) = [H(x)]^2, \quad \forall x \in B(x^*, \epsilon), \quad \epsilon > 0. \tag{4.24}$$

Hence

$$V(x) = U(x), \quad M(x) = [\Lambda(x)]^2, \quad \forall x \in B(x^*, \epsilon) > 0 \tag{4.25}$$

and from (4.23) we obtain, since $H(x) = U(x)\, \Lambda(x)\, U^T(x)$,

$$x(t) = x^k + \left[\sum_{i=1}^{n} \frac{\exp(-t\lambda_i^2(x^k)) - 1}{\lambda_i(x^k)}\, u_i(x^k)\, u_i^T(x^k)\right] g(x^k),\ x^k \in B(x^*, \epsilon),\ \epsilon > 0. \tag{4.26}$$

As may be easily proved

$$\lim_{\lambda_i(x^k)\to 0} \frac{\exp(-t\lambda_i^2(x^k)) - 1}{\lambda_i(x^k)} = 0 \tag{4.27}$$

whereas

$$\lim_{\lambda_i(x^k)\to 0} \frac{\exp(-t\lambda_i(x^k)) - 1}{\lambda_i(x^k)} = -t. \tag{4.28}$$

Note also that $J(x) \simeq [H(x)]^2$, $\forall x \in B(x^*, \epsilon)$, $\epsilon > 0$ is always positive semidefinite. Hence we can always in (4.25) let $t \to +\infty$, something which is admissible in (4.8) only when $H(x)$ is positive definite. Therefore system (4.19) is superior to system (4.6), as fewer conditions have to be imposed on the Hessian matrix. At this point we recall that it has been suggested [10] that a reduction in $\|g(x)\|$ rather than in $f(x)$ should be required at each step, as this might be useful in finding saddle points, and improve the performance of a certain minimization algorithm on flat functions. Therefore a combination of (4.7) and (4.20) could be an advantage for functions with saddle points or flat minima. Indeed, very satisfactory numerical results were obtained with an algorithm which uses (4.7) switching to (4.20) and requiring a reduction in $\|g(x)\|$ rather than in $f(x)$ when the objective function becomes insensitive to changes in the variables near the minimum.

For an extensive discussion of the properties of the above curvilinear search paths, as well as of those associated with the higher-order systems, see [11].

In the following section the general algorithm is stated and proved to be convergent.

5. The General Algorithm and its Convergence Properties

Let $p(x^k, t)$ be the curvilinear search path obtained by linearizing any of our systems of differential equations. As may be seen, $p\colon R^n \times R^{1+} \to R^n$ is of the form

$$p(x, t) = Q(x, t)\, g(x) \tag{5.1}$$

where $Q(x, t)$ is a negative definite matrix continuous in t, and satisfies the following conditions:

(i) $g^T(x)\, p^{[q]}(x, 0) < 0$ and $p^{[r]}(x, 0) = 0,\ r = 1, 2,\dots, q-1,$
$x \in \{x \mid g(x) \neq 0\}$;

(ii) $p(x, t) = 0,\ \forall t \in [0, +\infty),\ \forall x \in \{x \mid g(x) = 0\}$; and

(iii) $p(x, 0) = 0$.

If we now let

$$x(t) = x + p(x, t) \tag{5.2}$$

the following general algorithm may be stated for the minimization of $f(x)$, $f\colon R^n \to R^1$, over R^n.

The General Algorithm

Let $p = R^n \times R^{1+} \to R^n$ be a continuously differentiable mapping in t satisfying conditions (i)–(iii).

Step 0. Select an $x^0 \in R^n$ such that the level set $L_0 = \{x \mid f(x) \leqslant f(x^0), x \in R^n\}$ is compact.

Step 1. Set $k = 0$.

Step 2. Compute $g(x^k)$; if $g(x^k) = 0$, stop; else compute $p(x^k, t)$ as a function of t and go to step 3.

Step 3. Find the smallest non-negative scalar $t(x^k)$ minimizing $f(x(t))$ along $p(x^k, t)$.

Step 4. Set $x^{k+1} = x^k + p(x^k, t(x^k))$; set $k = k + 1$ and go to step 2. ▮

In order to establish the convergence properties of the above algorithm, we have to consider two cases separately, viz.

(a) $p(x, t)$ is jointly continuous in both its arguments; and

(b) $p(x, t)$ is continuous in t only.

The second case occurs whenever the derivatives of $f(x)$ of second or higher order are approximated numerically rather than computed analytically. A proof of convergence covering the first case, when $p(x, t) = [\exp(-tH(x)) - I]H^{-1}(x)g(x)$, may be found in [9]. In this paper we consider the second case. The following algorithm model and theorem are needed.

5.1. *Algorithm Model* (*Polak*, [12])

Let A be a set-valued search function mapping a closed subset T of a Banach space into the set of all non-empty subsets of T (which is written as $A\colon T \to 2^T$) and let c be a stop rule, $c\colon R \to T$.

Step 0. Compute an $x^0 \in T$.

Step 1. Set $k = 0$.

Step 2. Compute a point $y \in A(x^k)$.

Step 3. Set $x^{k+1} = y$.

Step 4. If $c(x^{k+1}) \geqslant c(x^k)$, stop; else set $k = k + 1$ and go to step 2. ∎

5.2 Theorem (Polak, [12]). *Suppose that*:

(1) *$c(x)$ is either continuous at all non-desirable* points $x \in T$, or else $c(x)$ is bounded below for $x \in T$; and*

(2) *for every $x \in T$ which is not desirable there exists an $\epsilon(x) > 0$ and a $\delta(x) < 0$ such that $c(x'') - c(x') \leqslant \delta(x) < 0$, $\forall x' \in T$ such that $\| x' - x \| \leqslant \epsilon(x)$ and $\forall x'' \in A(x')$.*

Then either the sequence $\{x^k\}$ constructed by the above algorithm model is finite and its last element but one is desirable, or else the sequence is infinite and every accumulation point of $\{x^k\}$ is desirable. ∎

The convergence properties of our general algorithm, when $p(x, t)$ is not continuous in x, may now be established. We assume that $p(x, t)$ satisfies conditions (i)-(iii) and that it is of the form

$$p(x, t) = [\exp(-tS(x)) - I]\, S^{-1}(x)\, T(x)\, g(x) \tag{5.3}$$

where $S(x)$ and $T(x)$ are symmetric matrices.

* A point x is a called desirable if $g(x) = 0$.

MAIN THEOREM. *Let $f(x)$ be a continuously differentiable function, $f: R^n \to R^1$. Assume that:*

(a) *there exist two constants μ and m, $0 \leqslant \mu < m, < +\infty$, such that*

$$0 \leqslant \mu \| y \|^2 \leqslant y^T H(x) y \leqslant m \| y \|^2, \quad \forall y \in R^n \qquad \textit{and} \qquad \forall x \in L_0;$$

(b) *$S(x)$ is positive definite with eigenvalues bounded from above and below by L and l respectively;*

(c) *$T(x)$ is positive semi-definite, $\forall x \in L_0$ and $\| T \| < b$, $0 < b < +\infty$;*

(d) *$g^T(x)\, T(x)\, g(x) \geqslant \epsilon \| g(x) \|^2$, $\epsilon > 0$, $\forall x \in L_0$; and*

(e) *$T(x)$ and $S(x)$ have a complete set of eigenvectors in common. Then the sequence $\{x^k\}$ constructed by the above general algorithm, with $p(x, t)$ given by (5.3), is either finite, terminating at a desirable point, or infinite and every accumulation point of $\{x^k\}$ is desirable.*

Proof. If we set $c(x) = f(x)$ and define a set-valued search function by

$$A(x) = x + p(x, t) \tag{5.4}$$

where t minimizes $f(x(t))$ along $p(x, t)$, then we see that our general algorithm is of the same type as Polak's algorithm model 5.1. Hence theorem 5.2 can be directly applied. Condition (1) of Polak's theorem is satisfied, since $f(x)$ is continuous. Let now x^k be a desirable point, i.e. $g(x^k) = 0$. Then by condition (ii) $p(x^k, t) = 0$, $\forall t \in [0, +\infty)$. Therefore $f(x^{k+1}) = f(x^k)$ and the algorithm terminates.

Now let $x \in R^n$ be non-desirable, i.e. $g(x) \neq 0$, and consider the mapping $\Delta: R^n \times R^{1+} \to R^1$ given by

$$\Delta(x, p(x, t)) = f(x + p(x, t)) - f(x). \tag{5.5}$$

By Taylor's formula for second-order expansion, we have

$$\begin{aligned} \Delta(x, p(x, t)) &= g^T(x)\, p(x, t) \\ &\quad + \int_0^1 (1-z)\, p^T(x, t)\, H(x + zp(x, t))\, p(x, t)\, dz, \; z \in (0, 1). \end{aligned} \tag{5.6}$$

Provided that $x + zp(x, t) \in L_0$, we have by assumption (a):

$$\Delta(x, p(x, t)) \leqslant g^T(x)\, p(x, t) + \frac{m}{2}\, p^T(x, t)\, p(x, t). \tag{5.7}$$

Now let $\lambda_n(x) \geqslant \lambda_{n-1}(x) \geqslant \cdots \geqslant \lambda_1(x)$ be the eigenvalues of $S(x)$ and $u_n(x)$,

$u_{n-1}(x), \ldots, u_1(x)$ the associated normalized eigenvectors. By assumption (b):

$$\lambda_n(x) \leqslant L, \qquad \lambda_1(x) \geqslant l, \qquad \forall x \in L_0 . \tag{5.8}$$

We also have

$$g^T(x)\, p(x, t) = g^T(x)[\exp(-tS(x)) - I]\, S^{-1}(x)\, T(x)\, g(x) \tag{5.9}$$

or

$$g^T(x)\, p(x, t) = g^T(x) \left[\sum_{i=1}^{n} \frac{\exp(-t\lambda_i(x)) - 1}{\lambda_i(x)}\, u_i(x)\, u_i^T(x) \right] T(x)\, g(x). \tag{5.10}$$

Expanding $g(x)$ into the eigenvectors of the symmetric matrix $S(x)$ we obtain

$$g(x) = \sum_{j=1}^{n} a_j(x)\, u_j(x). \tag{5.11}$$

From (5.10) and (5.11), and by assumption (e), we get

$$g^T(x)\, p(x, t) = \sum_{i=1}^{n} \frac{\exp(-t\lambda_i(x)) - 1}{\lambda_i(x)}\, a_i^2(x)\, \rho_i(x) \tag{5.12}$$

where $\rho_i(x) \geqslant 0$, $i = 1, 2, \ldots, n$, are the eigenvalues of $T(x)$. Now since since $a_i^2(x)\, \rho_i(x) \geqslant 0$ and $\exp(-t\lambda_i(x)) - 1/\lambda_i(x) < 0$, we have

$$g^T(x)\, p(x, t) \leqslant \frac{\exp(-t\lambda_n(x)) - 1}{\lambda_n(x)} \sum_{i=1}^{n} a_i^2(x)\, \rho_i(x) \tag{5.13}$$

and, since $g^T(x)\, T(x)\, g(x) = \sum_{i=1}^{n} a_i^2(x)\, \rho_i(x)$,

$$g^T(x)\, p(x, t) \leqslant \frac{\exp(-\lambda t_n(x)) - 1}{\lambda_n(x)}\, g^T(x)\, T(x)\, g(x). \tag{5.14}$$

Hence, by assumption (d),

$$g^T(x)\, p(x, t) \leqslant \frac{\exp(-t\lambda_n(x)) - 1}{\lambda_n(x)}\, \epsilon \, \| g(x) \|^2. \tag{5.15}$$

For the second term in (5.7) we have

$$\begin{aligned} p^T(x, t)\, p(x, t) &= g^T(x)\, T(x)\, [[\exp(-tS(x)) - I]\, S^{-1}(x)]^2\, T(x)\, g(x) \\ &\leqslant \left[\frac{\exp(-t\lambda_1(x)) - 1}{\lambda_1(x)} \right]^2 \rho_n^2(x)\, \| g(x) \|^2. \end{aligned} \tag{5.16}$$

Substituting (5.15) and (5.16) into (5.7) we obtain

$$\begin{aligned}\Delta(x, p(x, t)) &\leqslant \frac{m}{2}\left[\frac{\exp(-t\lambda_1(x)) - 1}{\lambda_1(x)}\right]^2 \rho_n{}^2(x) \, \| g(x)\|^2 \\ &\quad + \frac{\exp(-t\lambda_n(x)) - 1}{\lambda_n(x)} \, \epsilon \, \| g(x)\|^2 \qquad (5.17) \\ &\leqslant \frac{m}{2}\left[\frac{\exp(-t\lambda_1(x)) - 1}{\lambda_1(x)}\right]^2 b^2 \, \| g(x)\|^2 \\ &\quad + \frac{\exp(-t\lambda_1(x)) - 1}{\lambda_n(x)} \, \epsilon \, \| g(x)\|^2\end{aligned}$$

since $\exp(-t\lambda_n(x)) < \exp(-t\lambda_1(x))$; and, by assumption (c), $\rho_n = \| T(x)\| < b$.

The step size t minimizing the right-hand side of the above expression is given by

$$\frac{d\Delta(x, p(x, t))}{dt} = 0 \qquad (5.18)$$

from where we obtain

$$\tau = 1 - \frac{\epsilon\lambda_1{}^2(x)}{\lambda_n(x)\, mb^2}, \qquad \tau = e^{-t\lambda_1(x)}. \qquad (5.19)$$

Now, since by assumption (a), $l \leqslant \lambda_i(x) \leqslant L$, we obtain

$$\Delta(x, p(x, t) \leqslant -\frac{1}{2m}\left(\frac{l}{L}\right)^2 \left(\frac{\epsilon}{b}\right)^2 \| g(x)\|^2. \qquad (5.20)$$

By the continuity of $g(x)$, there exists an $\epsilon(x) > 0$ such that for all x such that $\| x' - x \| \leqslant \epsilon(x)$,

$$\| g(x')\|^2 \geqslant \tfrac{1}{2} \| g(x)\|^2. \qquad (5.21)$$

Therefore

$$f(x' + p(x', t)) - f(x') \leqslant -\frac{1}{4m}\left(\frac{l}{L}\right)^2 \left(\frac{\epsilon}{b}\right)^2 \| g(x)\|^2, \ x' \in B(x, \epsilon(x)). \qquad (5.22)$$

Recognizing that $x' + p(x', t) = x'' \in A(x')$ and letting

$$\delta(x) = -\frac{1}{4m}\left(\frac{l}{L}\right)^2 \left(\frac{\epsilon}{b}\right)^2 \| g(x)\|^2 < 0$$

we obtain

$$f(x' + p(x', t)) - f(x') \leqslant \delta(x) < 0, \quad \forall x' \in B(x, \epsilon(x)) \quad \text{and} \quad \forall x'' \in A(x').$$

Hence condition (2) of Polak's theorem 5.1 is satisfied and our theorem is proved, as accumulation points do exist, since the level set L_0 is compact. ▮

6. Numerical Results

Based upon our curvilinear search paths, a number of algorithms were derived and extensively tested in our thesis, [11]. The convergence properties of the algorithms were established by means of the main theorem of the previous section. In this paper, however, we shall present only one of the algorithms which is similar to that presented in [9], with the difference that it is not necessary to compute second derivatives. The curvilinear search path used in this algorithm is given by

$$p(x, t) = [\exp(-tH(x)) - I]\, H^{-1}(x)\, g(x). \tag{6.1}$$

Clearly the matrix of the second derivatives of $f(x)$ has to be computed analytically, and this is a feature that should be avoided, although it leads to extremely fast and stable algorithms. In order to approximate the Hessian matrix of the objective function using first-order information only, the following procedure is now used.

Let us assume that the function to be minimized is quadratic. Then for any two points x^i and x^{i+1} in R^n we must have

$$g(x^{i+1}) - g(x^i) = H(x^{i+1} - x^i) \tag{6.2}$$

where H is the constant Hessian matrix. The above equation evaluated at $n + 1$ distinct points x^i, $i = 0, 1,\ldots, n$, yields

$$\begin{aligned} g(x^1) - g(x^0) &= H(x^1 - x^0) \\ &\;\vdots \\ g(x^i) - g(x^{i-1}) &= H(x^i - x^{i-1}) \\ &\;\vdots \\ g(x^n) - g(x^{n-1}) &= H(x^n - x^{n-1}) \end{aligned} \tag{6.3}$$

or, in a matrix form,

$$\Delta G_n = H \Delta X_n \tag{6.4}$$

where

$$\Delta G_n = [\ldots, \Delta g_i ,\ldots] \tag{6.5}$$

$$\Delta X_n = [\ldots, \Delta x_i ,\ldots] \tag{6.6}$$

$$\Delta g_i = g(x^i) - g(x^{i-1}), \qquad i = 1, 2,\ldots, n \tag{6.7}$$

$$\Delta x_i = x^i - x^{i-1}, \qquad i = 1, 2,\ldots, n. \tag{6.8}$$

Assuming that the vectors Δx^i, $i = 1, 2,\ldots, n$ are linearly independent, we can invert ΔX_n, obtaining

$$H = \Delta G_n \Delta X_n^{-1}. \tag{6.9}$$

It is, however, desirable to carry out the inversion recursively, as new Δx_k and Δg_k are evaluated during successive steps of our iterative minimization process. We do this by defining $\Delta X_0 = \Delta G_0 = I$ (the identity matrix). Then at each iteration we successively replace corresponding columns of ΔX_k and ΔG_k by calculated values of Δx_k and Δg_k, i.e.

$$\Delta X_{k+1} = \Delta X_k + (\Delta x_k - \Delta X_k e_j)\, e_j^T \tag{6.10}$$

and

$$\Delta G_{k+1} = \Delta G_k + (\Delta g_k - \Delta G_k e_j)\, e_j^T \tag{6.11}$$

where e_j is a unit vector having unity as the j-th element, zeros elsewhere and $j = k + 1$. Now, using Sherman-Morrison's formula [2] we obtain

$$\Delta X_{k+1}^{-1} = \Delta X_k^{-1} - \frac{(\Delta X_k^{-1}\Delta x_k - e_j)e_j^T \Delta X_k^{-1}}{e_j^T \Delta X_k^{-1} \Delta x_k} \tag{6.12}$$

where of course $e_j^T \Delta X_k^{-1} \Delta x_k$ must differ from zero for ΔX_{k+1}^{-1} to exist. Hence if we denote by η_{k+1} the approximated Hessian matrix at x^{k+1}, we have

$$\eta_{k+1} = \eta_k - \frac{(\eta_k \Delta x_k - \Delta g_k)e_j^T \Delta X_k^{-1}}{e_j^T \Delta X_k^{-1} \Delta x_k}. \tag{6.13}$$

Obviously for a quadratic objective function $\eta_n = H$, i.e. *after n steps the minimum is reached.* Note also that, for a general function, η_k is not symmetric and therefore a symmetrization of the form

$$\eta_k = \tfrac{1}{2}[\eta_k + \eta_k^T] \tag{6.14}$$

is needed. The following algorithm may now be stated.

Step 0. Select an $x^0 \in R^n$ such that the level set L_0 is compact; set the tolerances l, L and ϵ.

Step 1. Set $k = 0, j = 1, \eta_0 = \Delta X_0 = I$.

Step 2. Set $p(x^0, t) = -tg_0$; go to step 4.

Step 3. Find the eigenvalues $\lambda_i(x^k)$ and associated normalized eigenvectors $u_i(x^k)$ of η_k, $i = 1, 2,\ldots, n$; if $l \leqslant |\lambda_i(x^k)| \leqslant L$, set $\lambda_i(x^k) = |\lambda_i(x^k)|$, $i = 1, 2,\ldots, n$, and

$$p(x^k, t) = \left[\sum_{i=1}^{n} \frac{\exp(-t\lambda_i(x^k)) - 1}{\lambda_i(x^k)}\, u_i(x^k)\, u_i^T(x^k)\right] g(x^k);$$

else set $x^0 = x^k$ and go to step 1; go to step 4.

Step 4. Find the smallest non-negative scalar $t(x^k)$ minimizing $f(x(t))$ along $p(x^k, t)$.

Step 5. Set $x^{k+1} = x^k + p(x^k, t(x^k))$; compute $g(x^{k+1})$; if $g(x^{k+1}) = 0$, stop; else compute the new approximatation η_{k+1} to the Hessian matrix from (6.12), (6.13) and (6.14); if $| e_j{}^T \Delta X_k^{-1} \Delta x_k | < \epsilon$, then set $x^0 = x^{k+1}$ and go to step 1; else go to step 6.

Step 6. Set $k = k + 1$; if $j = n$, reset $j = 1$; else set $j = j + 1$ and go to step 3. ▮

It is now not difficult to prove that the above algorithm complies with the conditions of the main theorem of the previous section. Indeed, the case $p(x, t) = -tg(x)$ is trivial, since then $S(x) = T(x) = I$, and conditions (b), (c) and (e) are clearly satisfied. Now when $p(x, t) = [\exp(-t\eta(x)) - I]\, \eta^{-1}(x)\, g(x)$, we have $S(x) = \eta(x)$ and $T(x) = I$, and conditions (c) and (e) are satisfied. Condition (b) is also satisfied, since in our curvilinear search path we replace $\lambda_i(x^k)$ by its absolute value, thus forcing the Hessian matrix to be positive definite. Note, finally that as a result of the condition in step 3 of the above algorithm, $|\lambda_i(x^k)|$ remains always bounded by l and L, $i = 1, 2, \ldots, n$. Hence the sequence $\{x^k\}$ constructed by the above algorithm converges to a desirable point. Moreover, if the objective function is quadratic, the minimum is reached after at most n steps.

TABLE 6.1

Numerical Results

	Fletcher–Powell			New algorithm		
Function and x^0	Function evaluations	Gradient evaluations	Total	Function evaluations	Gradient evaluations	Total
Rosenbrock's $(-1.2, 1)$	167	167	501	162	32	236
Helical $(-1, 0, 0)$	76	76	304	101	25	176
Quartic $(3, -1, 0, 1)$	80	80	400	137	43	309
Banana $(-3, -1, -3, -1)$	161	161	805	310	110	705

In order to reduce the number of function evaluations when executing the above algorithm, we carried out a unidimensional search only if the point $x^{k+1} = x^k - \eta^{-1}(x^k)\, g(x^k)$, obtained by letting $t \to +\infty$, did not improve the function value. The tolerances were set equal to $l = 10^{-7}$, $L = 10^7$ and $\epsilon = 10^{-24}$, and the execution of the algorithm was terminated if both $|f(x^{k+1}) - f(x^k)| < 10^{-8}$ and $\| g(x^{k+1}) \| < 10^{-4}$. The numerical results obtained on testing the algorithm on a number of test functions appear in Table 6.1, were they are also compared with the results obtained on running the IBM System/360 Scientific Subroutine Package version of Fletcher and Powell's algorithm. The comparison turns out to be in favour of the new algorithm, which also improves on the performance of the algorithm presented in [9]. Moreover, we feel that the numerically stable method derived by Kowalik and Ramakrishnan, [15], for the homogeneous model [10], may be applied to our iterative scheme as well and further improve its performance. Results in this direction will appear elsewhere.

7. Conclusions

Our differential-equation approach to function minimization makes it possible to:

(i) replace search rays by more general curvilinear paths;

(ii) dispense with the requirement of a positive definite Hessian matrix;

(iii) directly associate the eigensystem of the Hessian matrix of $f(x)$ or $\phi(x) = \frac{1}{2}g^T(x)\, g(x)$ with the function-minimization problem;

(iv) derive a number of algorithms which converge in a finite number of steps on quadratic functions and rapidly minimize general functions; and

(v) open a wide area for future research.

The algorithm presented in this paper is, of course, an improvement on the one derived and discussed in [9], in the sense that no second derivatives are needed analytically. However, an eigenvalue problem has still to be solved and this is time-consuming, especially for problems of high dimensionality. As we have pointed out in our thesis, a method of approximating the eigensystem of the Hessian matrix using only first-order information could be a possible subject for future research. Indeed, an approximation scheme for the eigenvalues and eigenvectors themselves is under development and results will soon appear.

In this paper, an approximate solution to our system of differential equations has been sought. The fact, however, that our problem

$$\min f(x), \qquad x \in R^n \tag{7.1}$$

may now be written in the form

$$\min f(x(t)),\ x^{[q]}(t) = F(x, t),\ x^{[r]}(0) = 0,\ r = 1, 2,..., q-1,\ x(0) = x^k \tag{7.2}$$

suggests the development of known numerical integration techniques for use in function minimization. In our thesis some results of this approach have also been presented and proved to be satisfactory. These results, also more recent ones and further developments, will appear in a future paper.

ACKNOWLEDGMENTS

I wish to express my sincere thanks to Professor D. H. Jacobson for his continuous encouragement and very useful discussions, during which many of the ideas so far developed were generated.

REFERENCES

1. D. M. HIMMELBLAU, "Applied Nonlinear Programming," McGraw-Hill, New York, 1972.
2. J. M. ORTEGA AND W. C. REINBOLDT, "Iterative Solution of Nonlinear Equations in Several Variables," Academic Press, New York/London, 1970.
3. W. C. DAVIDON, "Variable Metric Methods for Minimization," A.E.C. Research and Development Report No. ANL-5990 (Rev.), 1959.
4. C. G. BROYDEN, Quasi-Newton methods and their application to function minimization, *Math. Comp.* **21** (1967), 368.
5. R. R. FLETCHER AND M. J. D. POWELL, A rapidly convergent descent method for minimization, *Comput. J.* **6** (1963), 163.
6. R. FLETCHER, A new approach to variable metric algorithms, *Comput. J.* **13** (1970), 317.
7. L. VARGA, A class of methods for function minimization, *Studia Sci. Math. Hungar.* **4** (1969), 291.
8. J. N. DANIEL, "Convergent Step Sizes for Curvilinear Path Methods of Minimization," CNA-29 (1971), Center for Numerical Analysis, University of Texas at Austin.
9. C. A. BOTSARIS AND D. H. JACOBSON, A Newton-type curvilinear search method for optimization, *J. Math. Anal. Appl.* **54** (1976), 217–229.
10. D. H. JACOBSON, D. H. AND W. OKSMAN, An algorithm that minimizes homogeneous functions of N variables in $N + 2$ iterations and rapidly minimizes general functions, *J. Math. Anal. Appl.* **38**, 1972, p. 535.
11. C. A. BOTSARIS, "Differential Descent Methods for Function Minimization," Doctoral Thesis, University of the Witwatersrand, Johannesburg, South Africa, August 1975.
12. E. POLAK, "Computational Methods in Optimization: A Unified Approach," Academic Press, New York, 1971.
13. D. F. DAVIDENKO, On a new method of numerical solution of systems of nonlinear equations, *Dokl. Akad. Nauk. USSR (N.S.)* **88** (1953), 601.
14. C. A. BOTSARIS AND D. H. JACOBSON, Minimization of a nonlinear function in many variables; a differential equation approach. Paper presented at a "Symposium on Differential Equations held at The Council for Scientific and Industrial Research, Pretoria, April 1975."
15. J. S. KOWALIK AND K. G. RAMAKRISHNAN, A numerically stable optimization method based on a homogeneous function, submitted for publication.